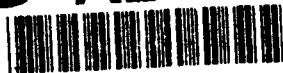


E 261 288

2

AD-A258 607

PL-TR-92-2017



**PL/GPD DMSP DATA ARCHIVING FROM VAX TO
EXABYTE TAPES**

**DTIC
ELECTS
OCT 5 1992**

**TASC
55 Walkers Brook Drive
Reading, Massachusetts 01867**

**David B. Lapadula
Maurice Ling**

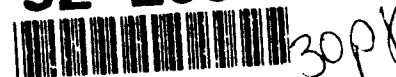
December 1991

Scientific Report No. 1

Approved for public release, distribution unlimited

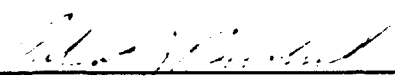
92 10 2 032

92-26387



**PHILLIPS LABORATORY
AIR FORCE SYSTEMS COMMAND
HANSCOM AIR FORCE BASE, MASSACHUSETTS 01731-5000**

"This technical report has been reviewed and is approved for publication"



ROBERT J. RAISTRICK
Contract Manager
Data Analysis Division



ROBERT E. MCINERNEY, Director
Data Analysis Division

This report has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify GL/IMA, Hanscom AFB, MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average one hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1991		3. REPORT TYPE AND DATES COVERED Scientific No. 1	
4. TITLE AND SUBTITLE PL/GPD DMSP Data Archiving from VAX to Exabyte Tapes				5. FUNDING NUMBERS F19625-91-C-0117 TASC No. J6397-211 PE 63214C PR 5321 TA 85 WU AB	
6. AUTHOR(S) David B. Lapadula Maurice Ling					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TASC 55 Walkers Brook Drive Reading, MA 01867				8. PERFORMING ORGANIZATION REPORT NUMBER TR-6397-2-1	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Phillips Laboratory Hanscom AFB, MA 01731-5000 Contract Manager: Robert Raistrick/LCY				10. SPONSORING/MONITORING AGENCY REPORT NUMBER PL-TR-92-2017	
11. SUPPLEMENTARY NOTES Includes User's Guide					
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report describes the archiving of Defense Meteorological Satellite Program (DMSP) data from VAX tapes to Exabyte tapes mounted on SPARCstations. This effort encompasses automatic creation of filenames for the data fields, transferral from one environment (VAX) to another (Unix), logistics for each step of the operation, devising a Data Dictionary file to provide the user information regarding the data format used for each file, writing a User's Guide for the archiving process, and training PL/GPD personnel how to archive and retrieve the data efficiently and accurately.					
14. SUBJECT TERMS DMSP SPARC stations VAX				15. NUMBER OF PAGES 30	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR		

TABLE OF CONTENTS

	Page
1. INTRODUCTION	1
2. BACKGROUND	2
3. ARCHIVING PLAN DEVELOPMENT	3
4. TECHNICAL DESCRIPTION OF ARCHIVING PROCESS	4
5. ARCHIVING BENCHMARKS	7
6. SUMMARY OF WORK COMPLETED	8
7. FUTURE EFFORTS	9
 ATTACHMENT A - User's Guide to the DMSP Archiving Process	 10
 APPENDIX A ARCTOOL PURPOSE	 24
APPENDIX B ARCTOOL FUNCTIONAL DESCRIPTION	25

DTC QUALITY INSPECTED 1

Accession For	
NTIS ARSI	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
Distribution	
Availability Codes	
Dist	Special
A-1	

1.

INTRODUCTION

The intent of this technical report is to present the reader with background information describing the efforts to archive the DMSP data located at PL/GPD, Hanscom AFB. This DMSP archiving process is part of Task 0002 under the Phenomenology Data Computations contract. The overall goal of this task is to establish procedures to archive all desired PL/GPD data files onto 8mm tapes, but efforts to-date have been directed towards archiving only the DMSP data.

The basic goal of this effort is to transfer the DMSP data from VAX-formatted 9-track tapes to Exabyte 8mm tapes mounted on SPARCstations running Unix. In order to successfully carry out this transfer, the following steps are necessary:

- Filenames automatically generated for each file, since the data currently have no associated filenames on the VAX tapes
- Files transferred from the VAX environment to the Unix environment
- Data copied to 8mm tapes in some sort of meaningful arrangement
- Logistics developed to catalog the data to facilitate accurate data retrieval from the 8mm tapes
- A Data Dictionary, containing the data format information for the files, created for each 8mm tape
- PL/GPD personnel trained to perform the archiving and retrieval in a timely and efficient manner.

Techniques for accomplishing all of the above steps have been successfully developed and are being implemented by personnel at PL/GPD for the F6 and F7 satellites. Although satellites F8, F9, and F10 (also DMSP) are currently still active, their backlog of data is also being archived in the same fashion as F6 and F7 data.

2.

BACKGROUND

PL/GPD has a large volume of data tapes stored off-site, as well as many on-site at Hanscom AFB. These tapes contain data from a myriad of experiments that were performed throughout the past few decades, and are accessed by the various Principal Investigators (PIs) in support of their research. Laboratory contractors maintain the data, retrieve the desired tapes from off-site when necessary, and run various tools against the data sets in support of the PIs.

This task was initiated under the Phenomenology Data Computations contract to support the archiving of this data to a media and environment that can be supported and accessed on-site at Hanscom by laboratory researchers and, eventually, the PIs themselves. Simply copying the data from one media/platform to another is not sufficient due to several factors:

- Conversion of data format at the bit level may be necessary (convert Cyber-based 60-bit words to a standard Unix 32- or 64-bit word format)
- A logistics process is required in order to determine which data has been transferred, which is in the process of relocation, and which data has been successfully moved from the original media/platform to the new one
- Retrieval methods for the new media may be significantly different than those of the old media.

3.

ARCHIVING PLAN DEVELOPMENT

Archiving problems arise from the fact that the PL Cyber is scheduled to be taken off-line in September 1992, but the majority of the PL/GPD data has been stored on Cyber-specific tapes, most of which are stored off-site in a warehouse in Lowell, MA. Because the older Cyber data is in a 60-bit word format, the data will have to be converted to a 16- or 32-bit word format, depending on the replacement mainframe selected. Assorted tools currently used on the Cyber to examine and process the data will have to be ported as well, but they are beyond the scope of this task.

Due to these conversion uncertainties, it was decided to first archive any data that is not stored as 60-bit words. DMSP data from satellites F6 through F10 (F6 and F7 are inactive; F8-F10 are still active) is such a source, and was selected because of the following reasons:

- The tapes are mainly on-site and readily available for archiving
- The formats are rigid and do not vary from tape to tape
- The formats are well-documented and easily understood
- There are hundreds of DMSP tapes that need to be archived, which is enough to serve as a testing ground for any of the archiving procedures developed
- The data has been (and is being) converted from Cyber 60-bit words to 16-bit words on VAX tapes, allowing archiving without bit-level conversions
- Data integrity is easily verified on both the VAX and the SPARCstation.

The target media for the DMSP data to be archived onto was defined as 8mm tapes to be used with Exabyte tape drives on SPARCstations. Although both 4mm DAT and 8mm tapes were originally considered, 8mm tapes had more favorable reasons to justify their usage. These factors included its high storage density (2GB or 5GB), availability (8mm is to be used throughout the rest of the MSX project), portability (tape is the size of a cigarette package), data format standardization (only 2 possible data storage formats), and cost (under \$20/tape, with the drives costing only several thousand dollars each). Although the 8mm drives will be connected to SPARCstations, this is merely for storage and retrieval activities. The data will still have to be processed on a separate platform containing the analysis tools.

4. TECHNICAL DESCRIPTION OF ARCHIVING PROCESS

The process of transferring DMSP data from 9-track tapes to 8mm Exabyte tapes occurs in three separate steps:

- (1) The tapes are downloaded onto the VAX disk;
- (2) The data is transferred to the SPARCstation; and
- (3) The data is stored on 8mm tape media.

The following paragraphs describe, in technical detail, the files and programs necessary for each step of the process. A system-oriented approach was used, which utilized the ability of the VMS and Unix systems to process command files and scripts.

The first step involved enhancing the VAX program *tapecopy* to write a Unix-compatible binary file when it transferred the data files from the 9 track tape to the VAX disk. This modification was necessary due to fact that the original *tapecopy* produced variable format binary files which contained extraneous data at the beginning and end of each file. The resulting program which produced these Unix-compatible files is named *utapecopy*. To facilitate the operation of *utapecopy*, a command file named *detool.com* was created. This file contains the necessary output data file names and inputs for *utapecopy* to successfully download one 9-track tape. The files are named FILE ##.DAT, where ## is the number of the file on the tape. These files will be automatically renamed after transferring them to the SPARCstation. In order to automate the archiving process, the command file *tapebatch.com* was created. This command file contains commands to check for any available tape drives and to mount a tape with a user specified label. Following the mounting command, the command file executes *detool.com* to download the data on the disk. Thus, by submitting *tapebatch.com* to the VAX *tapejob* command, the data files can be downloaded without any user supervision. At this point, it is suggested that the archiver minimizes the wait time by calling the VAX operator (x2642) to activate the batch job.

The second step of data transfer entails the use of FTP (File Transfer Protocol), which involves noninteractively logging onto the SPARCstation and sending the files from the VAX to the SPARCstation. The user must ensure that the FTP transfer occurs in

binary mode to avoid corrupting the data. After transferring the files, the user should execute the *autoname* program to rename the files according to the date in each file. The format of the file name will be YYMMMDD.dat, where YY is the last two digits of the year, MMM is the three letter abbreviation for the month, and DD is the day of the month. If there happen to be multiple files for a day, a suffix of ## will be added, where ## is a two digit number corresponding to the number of extra files.

The third and final step of the data transfer process requires the use of the program *arctool* and its utility files, *copyformat* and *lastfile*. *Arctool* utilizes the Unix programs *mt* and *tar* to control access to the Exabyte drive, which is represented as the variable TAPE, defined as "/dev/rst1." The variable NTAPE, defined as "/dev/nrst1," is used for non-re-winding device access. These variables can be modified by editing the *arctool* file in the bin directory.

The *arctool* files are organized into different directories (archive, bin, formats, and catalogs) which are in the user's home directory. The data files transferred from the VAX should be placed in the "archive" directory, where *arctool* will read them and transfer them onto an Exabyte tape. The "bin" directory contains *arctool* and its utility files, and must be in the path specified in the user's *.cshrc* file. *DMSP.format*, the data dictionary or format specification file, resides in the "format" directory, and the data file names on each Exabyte tape are contained in separate catalog files in the "catalogs" directory. When the user is archiving the data, the catalog file name should be specified.

When starting a new tape, the *addformat* program copies the *DMSP.format* file onto the tape as the first file and adds its name to the specified catalog file. This format file should be the first file on each Exabyte tape. The data files, separated by file markers, are then added to the tape. When monitoring the space taken up by files on the tapes, one must take into account the size of the file markers. For the Exabyte 8200, each file marker occupies approximately 3 megabytes of storage space.

The *lastfile* program gives the user the ability to find the file number of the last file in the catalog file of the tape. The user can then use *arctool* to forward the tape to the end of the data and add more files starting at that point in the tape.

To retrieve files from the Exabyte tape, *arctool* allows retrieval by individual file numbers, sequential file numbers, or an arbitrary number of specified file numbers,

which must be in ascending numerical order. File numbers can be obtained by using the *findfile* program and specifying the name or portion of the name of the data file to retrieve and the name of the catalog of the Exabyte tape.

In order to access the VAX computer through the workstation, it is necessary to use the terminal program, *xterm*, as the main terminal window. The other OpenWindows shells and command windows did not cooperate with the VT100 codes of the VAX/VMS operating system. Since *xterm* was not installed on the archiving workstation, it was necessary to copy it from another workstation, and to place it in the user's bin directory along with *arctool* and its utilities.

5.

ARCHIVING BENCHMARKS

According to initial tests, it should be possible to archive approximately 6 months of DMSP data onto one 8mm 2GB Exabyte tape. The amount of time it takes to actually perform the entire data archive transfer from 9-track to 8mm varies according to the awareness of the tape mount personnel, the load on the system, etc. However, it is suggested that one 8mm tape can be successfully filled each workday; thus, one year's worth of data should be archivable within two days, assuming no interruptions occur in the process. The current listings of data from satellites F6 through F9 (there was no listing for F10) reveal 18 years worth of data presently ready for archival, with more data flowing from F8-F10 at this time. This corresponds to a level of effort of two man-months to perform the archiving process. It should be understood, however, that this figure could be significantly reduced by performing the archiving simultaneously on several machines, and also by performing individual up- and down-loading in parallel on the same terminal.

6.

SUMMARY OF WORK COMPLETED

The archiving and data transfer process was presented to PL/GPD in four separate ways, as listed below.

- (1) The "User's Guide to the DMSP Archiving Process" (Attachment A) was written as a step by step user-oriented description of the data transfer process. After many revisions, this document was delivered to those involved in the actual data transfer and archiving process.
- (2) An official documentation form, AFSC Form 1546, was completed to document the functional, hardware, and software requirements of arctool, the core tool that was developed to archive data to and retrieve data from Exabyte tapes mounted on SPARCstations.
- (3) A hands-on tutorial was given to the technicians who were responsible for undertaking the DMSP archiving task. This tutorial covered most of the material described in the "User's Guide to the DMSP Archiving Process." The archiving accounts and disk space were set up and fully operational for the data transfer and archiving task.
- (4) This technical report was written to summarize the efforts and information involved in the DMSP archiving task.

7.

FUTURE EFFORTS

There are several options currently being considered for the next step of this data archiving subtask. Among them are:

- Archiving off-site non-Cyber data (no bit-level conversions necessary)
- Archiving all of the other on-site data (most Cyber-specific), possibly requiring a conversion effort to make the 60-bit words readable to the Unix servers
- Concentrate on PL/GPO data archiving, which means a much more vigorous interconnection between the archived data and forthcoming MSX data.

The contract monitor will provide directions for which path may be undertaken in the near future. A major factor which should also be considered is that the Cyber is scheduled to be taken off-line at the end of this government fiscal year (late 1992); thus, any old Cyber-formatted data (7-track or 800 cpi tapes) will not be easily read or manipulated after that point. A possible solution that may be examined at the appropriate time is using the newly-installed Unix Migration System at Hanscom.

ATTACHMENT A
USER'S GUIDE TO THE DMSP ARCHIVING PROCESS

Preface

The following is a compilation of instructions for transferring DMSP data from 9-track VAX tapes to 8mm Exabyte tapes and for retrieving data from the Exabyte tapes. Before doing the archiving or retrieval, the environment must be set up correctly. The first section, **User Environment Setup**, is for configuring both the VAX and SPARCstation accounts for optimum utilization of the archiving software. If the accounts are already set up, the user can proceed to the second section, **Archiving Steps**, which is a step-by-step description of the archiving process. The third section, **Retrieval Steps**, contains a step-by-step description of retrieving data which has been already archived onto 8mm tapes using the procedure described in the **Archiving Steps** section. The fourth section, **Helpful Hints**, contains suggestions which the user should keep in mind while archiving and retrieving data. The last page is a brief summary of each step of the archiving process which the user can use as a quick reference. These instructions assume that the user has rudimentary knowledge of the Unix and VAX operating systems and text editors.

On a technical note, the archived data are stored in a Least Significant Byte (LSB) first format, due to the way the VAX stores binary data. This byte ordering poses no compromise to data integrity, but should be noted when manipulating the data after retrieval.

User Environment Setup

On the VAX...

1. Create two directories — one to place the tape data, and the other to place executables and command files. Typically, the data directory will need to be created on another disk. Make sure there are at least 300000 to 400000 blocks of space in the data directory by typing 'show quota' while you are in your data directory. This amount of space should be sufficient for downloading data from one entire 9-track tape.

Example:

```
$ set default lcy$disk:[lc.ling]
```

```
$ show quota
```

```
User [LCY, LING] has 200757 blocks used, 99243 available, of
300000 authorized and permitted overdraft of 100 blocks
```

2. Define the following in your **login.com** file.

ARCHIVE — the directory where the tape data is to be placed

BIN — the directory holding the executables and command files

CD — "set default"

Example:

Your login.com file should have the following lines in it:

```
$ DEFINE ARCHIVE $1$DUA0:[DAILY_SCRATCH.ARCHIVE]
```

```
$ DEFINE BIN USR$DSK4:[LCY.LING.BIN]
```

```
$ CD == "set default"
```

You need to modify the BIN directory name to be your own BIN directory.

3. Copy the following files into the BIN directory:

```
USR$DSK4:[LCY.LING.BIN]TAPEBATCH.COM
```

```
USR$DSK4:[LCY.LING.BIN]DETOOL.COM
```

```
USR$DSK4:[LCY.LING.BIN]UTAPECOPY.EXE
```


On the SPARCstation...

1. Using **mkdir**, create the following directories in your home directory:
 - bin Contains the executable files
 - archive Contains the data to be archived
 - catalogs Contains the catalog files for the Exabyte Tapes
 - formats Contains the DMSP.format file
2. Copy **arctool**, **autoname**, **addformat**, **findfile**, and **lastfile** into the bin directory. If **xterm** is not installed on the system, copy **xterm** from `~ling/bin` or any other available source.
3. Add the following alias to your **.cshrc** file.

```
alias vax 'telnet 146.153.100.9'
```
4. Add `~/bin` to the path specified in the **.cshrc** file.
5. Make sure there is enough space on the disk to store the tape data by typing `'df'`. There should be at least 180 megabytes available on your disk. The more disk space you have, the more 9 track tapes you can archive at one time.
6. Make sure that you have a **.Xdefaults** file and a **.openwin-menu** file in your home directory. (e.g., `ls -a`). If not, copy these from the directory `~ling`. Log out and log in again to activate any changes.
7. In the **.login** file, set the environment variable, **ARCHIVE**, to be the archive directory.

Example:

```
setenv ARCHIVE /home/msxd/ling/archive
```

Archiving Steps

1. Obtain the listing of DMSP tapes to be transferred to Exabyte tapes. You will use this listing to keep track of the data.
2. Log onto **msxd**.
3. Select **xterm** from the system **programs** popup menu by clicking on the right mouse button on the background or, if **xterm** is not in the menu, click your mouse in the console window and type '**xterm&**'. Wait for a terminal window to pop up. All the commands from now on will be entered in this window.
4. Type '**vax**' to log onto the VAX.
5. Edit the file **tapebatch.com**. (e.g., edit **bin:tapebatch.com**) Change the tape name in the line containing the **MOUNT** command to the name of the tape which you want to mount.

Example:

Modify the following line in the file **tapebatch.com**:

```
$ MOUNT/for 'drv1' "M11083" TAPE1! request to load tape
```

Where M11083 is the name of the tape to load. The tape name is specified on the tape list.

6. Submit your job using **TAPEJOB**.

Example:

```
$ tapejob
```

```
_Batch job name: bin:tapebatch.com
```

```
Is your job name BIN:TAPEBATCH.COM?: y
```

```
Do you wish to include additional parameters? Y/N - : n
```

```
_Number of drives needed — Min. 1, Max. 3 : 1
```

```
You asked for 1 tape drives. Is this correct?: y
```

```
_Enter approximate CPU time in minutes: 30
```

```
You estimated for 30 CPU minutes. Is this correct?: y
```

You are logged onto GL9000, select the queue you wish to use.

1.) SYS\$MAG_TAPES_AFGLSC

2.) SYS\$MAG_TAPES_GL9000

Enter the number of the queue you wish to use. — : 1

Job BIN:TAPEBATCH.COM_1_td_30_m (queue ...)

Your Job has been submitted to the ... queue

See VMS HELP for a description of TAPEJOB. This command will submit the job to the SYS\$MAG_TAPES queue. The files will be extracted to the directory specified by ARCHIVE. The log file which contains the input and output listing from the job will be written to a file corresponding to the name of the job. It may take the operator a while to mount the tape, so be patient. You can check the process of the transfer by looking at the log file (e.g., type bin:<log filename>). The user is suggested to call the VAX operator at X2642 to promptly activate the batch job.

7. After the data is extracted to the VAX disk, type 'dir/size' to check the sizes of the files on the disk. Each file is named file##.dat, where ## is the relative number of the file on the tape. The file with the greatest number should be of size 0. Make a note of any abnormalities on the list of tapes.
8. Make a check mark to the left of the tape name on the list to indicate completion of 9-track tape to VAX transfer.
9. Use ftp in binary mode to transfer the files to the Unix system.

Example:

\$ cd archive

\$ ftp msxd

PL9000.PLH.AF.MIL MultiNet FTP user process 3.0(102)

Connection opened (Assuming 8-bit connections)

<MSXD FTP server (SunOS 4.1) ready.

MSXD.PLH.AF.MIL>user

Foreign username: **ling**

<Password required for ling.

Password:

<User ling logged in.

```

MSXD.PLH.AF.MIL>bin
Type: Image, Structure: File, Mode: Stream
MSXD.PLH.AF.MIL>cd archive
<CWD command successful.
MSXD.PLH.AF.MIL>mput file*
...
MSXD.PLH.AF.MIL>quit
<Goodbye.

```

10. Make a check mark to the right of the tape name on the list to indicate completion of the VAX to Unix transfer. You can now either continue to step 11 or delete the files on the VAX and prepare and submit the next batch job. It is good to check the Unix system to ensure the success of data transfer before deleting the files.
11. Logout of the VAX to return to the SPARCstation.
12. Set your directory to where the data resides.

Example:

```
% cd archive
```

13. Type '**autoname**' to automatically rename the files in the format **YYMMMDD.dat**, where **YY** is the last two digits of the year, **MMM** is the three digit abbreviation for the month, and **DD** is the day of the month. If there are multiple files for one day, the suffix '**_##**' will be added, where **##** is a sequential number corresponding to the number of extra days. Files of size zero will not be renamed, and should be deleted from the disk. Any extra files will be indicated in the output of **autoname**. If a file does not contain any date information, it will be considered an extra file, which will be named after the previous file with the appropriate suffix. The user should verify any such extra files with the description in the list of tapes.
 14. Type '**ls -l | more**' to obtain the size of the disk space used by the files. The total size, of data, in kilobytes, is at the top of the listing. Record this amount next to the tape or group of tapes on the tape list. Add this total to your running total for the Exabyte tape.
- * File markers are approximately 3 megabytes for the Exabyte 8200, so remember to add 3 megabytes for each file to the running total.

- * We suggest to put six months of data (six 9-track tapes) on each Exabyte Tape.
- 15. Place the correct Exabyte tape into the drive if the tape is not already loaded. If this is a new tape, proceed to step 16a. Otherwise, if you want to add files to the end of the data on the tape, continue to step 16b. If the tape is already in the drive and it is at the end of the data on the tape, advance to step 17.
- 16a. For a new tape, type '**addformat catalog_file**' to add the format file, DMSP.format to the beginning of the tape. See step 17 for a description of catalog_file. Proceed to step 17.
- 16b. If the tape already has data on it and you are uncertain that the tape is not already at the end of the data, you may want to advance the tape to the end of the data before you add the new data to the tape. The -A feature is currently disabled, since it is incompatible with the Exabyte drive. Instead, you can simply use the **lastfile** command to find the last file on the tape.

Example:

```
% lastfile tape_catalog  
344 tape_catalog
```

From this example, we see that there are 344 files on the tape catalog file, tape_catalog. This assumes that the tape listing contains one file name per line. Note that the files on the tape are numbered starting from 0. File number 0 is the DMSP.format file. For this example, to advance to the end of the tape, first rewind the tape, and then advance the tape:

```
% arctool -R  
% arctool → 344
```

Now, continue to step 17.

- 17. Use '**arctool -a files -F catalog_file**' to archive the data to tape. If this is a new tape, put the format file, DMSP.format, before the *files* on the command line. A file, catalog_file, containing the listing of the files being archived onto the tape will be created if it is a new catalog. If catalog_file exists already, the file will be added to at the end, and the old catalog list will be renamed to be

catalog_file.bak. This feature is useful if the user has accidentally specified the wrong catalog file. If this mistake occurs, the user can simply compare both files, and use a text editor to add the newly added file names to the correct catalog file.

All the catalog files are automatically located in a separate directory called ~/catalogs. Before adding files to the tape, it is a good idea to confirm that you are at the right position and file number on the tape. To determine the current file position of the tape drive, type: **arctool -s**.

Examples:

* For archiving 83jan01.dat, 83feb01.dat, and 83mar01.dat:

% arctool -a 83jan01 83feb01 83mar01 -F first_catalog

* For archiving all files beginning with 83 in the file name:

% arctool -a 83* -F 1983catalog

* For archiving all files beginning with 83jan, 83feb, and 83mar:

% arctool -a 83jan* 83feb* 83mar* -F 1983catalog

The arctool functions are described when you type **arctool** with no arguments:

MSXSER% arctool

Usage: arctool [-flag option(s)] ...

[-t opt]	No opt: List current file all: List all files end: List from current to end
[-A file1 file2 ...]	Archive named files to tape end
[-a file1 file2 ...]	Archive named files at current position
[-F cfile]	Add listing from -a, -A, or -t to cfile
[-c args]	No args: Extract current file n: Extract file number n * n1 n2: Extract from n1 to n2
[-m num1 num2 ...]	Extract multiple single files NOTE: Files must be in numerical order
[-+ num]	Advance tape num EOF markers
[-= num]	Reverse tape num files
[-s]	Status of tape
[-R]	Rewind tape
[-E]	Erase tape

* NOTE: Files numbering starts with 0

18. Label the tape appropriately. (eg., Jan - Jun 1983 —DMSP)

Retrieval Steps

1. Obtain the 8mm tape which contains the data to be retrieved and make sure it is *write protected* by checking the tab located on top of the tape. Insert the tape into the drive.
2. Find the file number of the file you want to retrieve using the utility, **findfile**:

Example:

```
% findfile jun30 mar10 dec21 -F file_listing
```

```
file number for jun30 is: 240
```

```
file number for mar10 is: 100
```

```
file number for dec21 is: 353
```

This is the file number which you use in step 4 below.

3. Change your current directory to be where you want the data.
4. Use 'arctool -c' or 'arctool -m' to extract the data from the tape. For multiple files, make sure the file numbers are listed in ascending order.

Examples:

* To extract the current file to disk:

```
% arctool -c
```

* To extract file number 50 :

```
% arctool -c 50
```

* To extract file numbers 50 to 79:

```
% arctool -c 50 79
```

* To extract file numbers 58, 87, 133, 256, and 300:

```
% arctool -m 58 87 133 256 300
```

Helpful Hints

1. Since it takes a long time to run some of these commands, you may want to run it while you are logged off. Unix provides a convenient program, `at`, to allow you to run commands at a scheduled time.

Example: To run the commandfile at 5:00pm,
% at 1700 commandfile

Here, `commandfile` is an optional file that contains the commands which you want to execute. If you do not put a `commandfile` after `at`, then you can manually enter in the commands you desire to execute. To exit `at`, type `[control]d` on a new line. For more information, consult the man pages for `at`. (e.g., `man at`) After the command file is executed, the system will send you mail containing the output of the commands.

2. The '`arctool -a files -F catalog_file`' archive option may be used consecutively without rewinding the tape. The files will continue to be added to the end of the data on the tape. Be sure to specify the correct catalog name.
3. As each Exabyte is filled, make sure you keep track of the total amount of data being put on the tape. This is easily done by doing a '`ls -l | more`' command in the archive directory. This will provide the total disk space contained in the directory, and the size of each file.
4. Use '`delete/noconfirm`' command to delete the files from the VAX/VMS disk after you transfer them to the SPARCstation. On the SPARCstation, use the '`rm`' command to remove the data files after you archive them to tape.
5. After transferring the data files from the VAX to the Unix system, you may want to delete the data on the VAX system and prepare and submit the next tape job. This will allow you to concurrently archive files on the Unix system while downloading files from tape on the VAX system.
6. When the tape is near capacity, make sure the data to be archived will not exceed the tape capacity. Due to the tape block size, the data archived will take up slightly more space on tape than on disk. In addition, each file has a

file marker which takes up 2.7 megabytes of space for the 2.3 gigabyte Exabyte 8200. Thus, be more conservative and do not try to archive too much data on the tape. The suggested amount to put on each tape is about 6 months of DMSP data.

7. When each Exabyte tape is completed, make sure to *write protect* it by sliding the tab on the top of the tape to the correct position. You may want to get a listing of all the files again to confirm that all the files were archived successfully. This can be accomplished by the following command:

```
arctool -t all -F catalog_filename2
```

You should make this catalog file name different than the one used when you were archiving. This way, you can compare both files to ensure consistency. Note, however, that the format of the catalog file names that are created while archiving to tape will be different than the format of the catalog file names that are created when you do a listing.

8. There is nothing special about catalog files. Catalog files are simply text files which can be edited if needed. Remember that if you forget to specify the catalog file, the listing goes into the file, `file_list`. This file should be deleted or renamed to avoid future confusion. A back up of the previous catalog file (with a `.bak` extension) is made whenever files are added to the list.

Archiving Process Summary

1. Login to msxd or host computer and start an **xterm**.
2. Type "**vax**" to log onto the VAX.
3. Edit **bin:tapebatch.com**
4. **tapejob** [call VAX operator to execute batch job]
5. **ftp hostname** [msxser, msxd, etc..]
[login sequence]
bin
cd archive_directory [should be ~/archive]
mput files
quit

[After transferring the data files to the SPARCstation, you may want to delete the data files on the VAX, and then start another tape job before continuing]
6. **logout** [from VAX to return to SPARCstation]
7. **autoname**

[If continuing archive process (tape is in drive already and tape is at the end of the tape data) go to step 12. If starting a new tape, go to step 8. If tape is not at end of data, go to step 9.]
8. **addformat catalog_file** [go to step 12]
9. **lastfile catalog_file** [output is num]
10. put correct tape in drive
11. **arctool -+ num**
12. **arctool -s** [ensure that tape is at the correct file]

13. **arctool -a *files* -F *catalog_file***
14. **ls -l | more** [record total size]
15. delete *files* on SPARCstation
16. delete *files* on VAX [reiterate from step 3 for the next tape]

APPENDIX A

ARCTOOL PURPOSE

The purpose of **arctool** is to facilitate archiving and retrieving files to and from 8 mm Exabyte magnetic tapes. This program was written to be the Unix part of the process of transferring data from DMSP 9-track data tapes in the VAX format to 8 mm Exabyte tapes mounted on SPARCstations. Other useful utilities for the Unix part of the process include **autoname**, which renames the data files according to the date in the header of the file, **addformat**, which places the format specification file onto the beginning of the tape, **findfile**, which outputs the file number of a file name in a catalog file, and **lastfile**, which outputs the file number of the end of the tape. On the VAX side of the process, UTAPECOPY, a modified version of TAPECOPY, is used to copy the 9-track tapes onto disk. There are also useful command files called DETOOL.COM and TAPEBATCH.COM which automate the downloading process. Please refer to the diagram below for an overview of the tape archiving process.

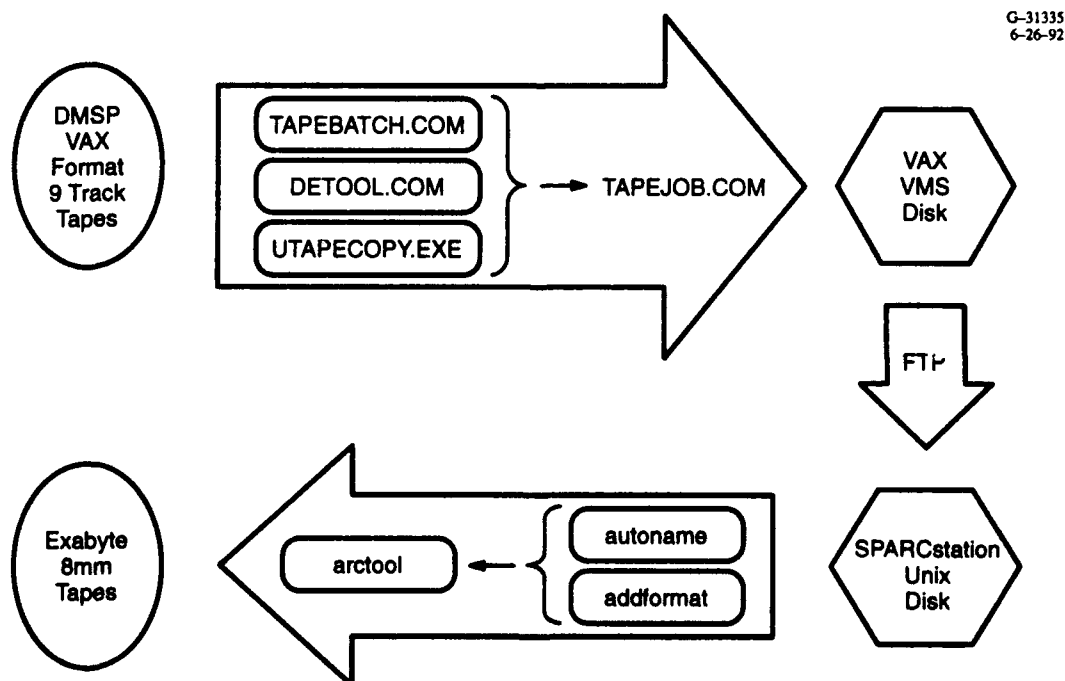


Figure 1 Archiving Overview

APPENDIX B

ARCTOOL FUNCTIONAL DESCRIPTION

Arctool is written using the Unix C shell (csh) script language. The files are stored on tape as individual tar files, separated by file markers. The basis of this script is the utilization of two Unix utilities, tar and mt. The input to the program is based on command line parameters. The generic form of input is: **arctool** *-flag parameter(s)* ... Where *flag* is a character that represents a specific function of arctool, and *parameter(s)* may be options, file numbers, or file names depending on the function desired. The flags and their parameters in the following list can be obtained by typing **arctool** at the Unix command line prompt.

Flags	Parameters	Description
-t	[None] all end	List current file List all files List current file to end of tape
-A	file1 file2 ...	Archive file names to tape end [Currently disabled due to incompatibility with Exabyte Drive]
-a	file1 file2 ...	Archive file names starting from current position
-F	catalog file	Specifies the output of the -a and -t commands to be written to or appended to the catalog file. If this flag is omitted the default file will be 'file_list'
-c	[None] <i>n</i> <i>n1 n2</i>	Extract current file from tape to disk Extract file number <i>n</i> from tape Extract file numbers <i>n1</i> to <i>n2</i> from tape
-m	<i>n1 n2 n3 ... n</i>	Extract multiple file numbers File must be in numerical order
-+	<i>n</i>	Advance tape <i>n</i> files
-=	<i>n</i>	Reverse tape <i>n</i> files
-s		Status of tape drive
-R		Rewind tape
-E		Erase tape

When file names are needed as parameters, wildcards used by the Unix C shell such as * and ? may be used. The first file, file number 0, should always be the data format specification file. The output of the program is written to the standard output in addition to a user specified text file when applicable. While the user is archiving data or obtaining a listing of the files on the tape, the output consists of the file names and their sizes.